

QNR data is arriving – first impressions

This will be an ongoing post for the next few weeks as data from my main research questionnaire begins to arrive in my e-mail InBox.

Getting co-operation to promote and deploy the QNR has been a more challenging part of the project than I was anticipating, but at least now (late Feb / early March 2016) some data is arriving.

The QNR has been deployed to a closed e-mail list of students registered with the university's Wellbeing Service and also deployed to the wider university community through a link on the students' UniHub home page.

The rationale for this dual deployment is that in the first instance, QNR returns are required from students indicating dyslexic learning differences in order that a baseline dataset can be constructed for creating the comparator profile for dyslexia. To date (3rd March 2016) 12 responses have been received from a total distribution a week or so ago to approximately 550 e-mail addressees which is disappointing, but I should remain patient.

Secondly, the deployment to the wider university community is to gain responses from any students with a view to identifying those who are indicating a profile aligned with dyslexia but who are disclosing in their responses that they have no learning challenges identified. So far, 15 QNR replies have been received over a similar time period, which is equally disappointing given the 30,000+ students registered at Middlesex University.

I have requested a second deployment through the Wellbeing Service and a second deployment to the wider student community through UniHub is scheduled anyway for next week.

Data analysis so far:

Including data from myself (accurately sent – is this ethical to include?) and from my son, who is a bona-fide student, there are 16 datasets so far in the research group ND – that is, students indicating no dyslexic learning difference.

In applying the Dyslexia Index calculator I have devised to QNR items 3.01 – 3.20, these 16 datasets produced Dyslexia Index Values ranging from the lowest at 141.85 to the highest of 936.05 (out of 1000).

BlogPost #17: QNR DATA IS ARRIVING – FIRST IMPRESSIONS

For the research group DI – students who indicated a dyslexic learning difference on their QNR response, the 12 datasets in this group so far produced Dyslexia Index Values ranging from 444.07 to 813.27.

From these datasets so far I have had to determine a ‘cut-off point’ along the range of values for the Dyslexia Index that marks the boundary of the evaluator ascribing an indication of alignment to a dyslexic profile, or not.

Working with the datasets for research group DI the mean value was calculated (646.79) together with the Standard Error of the Mean for this group (39.77).

Given that for a normal distribution it is accepted that 99% of sample means are likely to be within 3 standard deviations of the population mean, these calculations suggest a lower limit of 527.44 for the Dyslexia Index as the boundary for which it would be unlikely for someone with dyslexia, according to the criteria being used, to occur. [$527.44 = 646.79 - 3 \times 39.77$].

So this Boundary Value of a Dyslexia Index of 527 is tentatively set as the critical value above which a respondent may be considered to be exhibiting a profile that is more aligned with dyslexia than not.

Based on this, I can now create the mean data set for research group DI for creating the LoC Profiles which will be based, for now at least, on 8 of the 11 QNR response datasets collected to date. Only 8 are being used rather than the complete set of 12 because 3 QNR respondents' Dyslexia Index Values fell below the Boundary Value of 527. This in itself is quite interesting and is indicating that according to the criteria I have used for determining a dyslexic profile, 3 respondents who have declared their dyslexia don't have profiles aligned with dyslexia. I will need to explore this anomaly later.

When the Boundary Value is applied to the current dataset for research group ND it is appearing to be quite a promising discriminator as values for the Dyslexia Index in this group so far either fall below 346 or are above 543. With the Boundary Point set at 527 this now may be indicating my desired outcome of identifying respondents who are aligned with the dyslexic profile but who have not declared dyslexia

It is early days though, although this preliminary analysis looks promising.

The next step is to establish the mean profiles for research DI and ND based on this preliminary data that has arrived so far.

Establishing mean profiles:

This is going to be trickier than I had foreseen.

I am much aware that there are ‘lies, damned lies and statistics’ and it is entirely possibly to manipulate statistics to indicate a desirable result. To avoid this, and to equally avoid damning criticism resulting from the later scrutiny of my data analysis I am thinking very carefully indeed about the most appropriate way to create the baseline profiles for research groups DI and ND. Clearly I need to use the data that is arriving so far to achieve this but as would be expected in ‘real data collection’ there is plenty of variation in the datapoints collected in each of the datasets so far. I am also minded that by calculating a simple mean value, although this has the effect of diluting extreme values, this process of dilution increases as more data is included. A working example, I think, of the idea of ‘regression to the mean’!

Now for datasets in research group DI, the process is perhaps more straightforward and the initial process that I tried out is described above. However, I have reflected on using simple mean values to create the baseline comparator profile and decided that this can be improved to provide a more realistic representation of the most ‘typical’ student with dyslexia – which, after all, is the essential point for comparison.

So the first modification has been to use the mean values for the datasets that have Dyslexia Indices that are within 3 Standard Errors of the overall mean.

In order to be consistent, the same approach is applied to the datasets for research group ND and these two baseline comparator profiles are now used.

UPDATE: 10th March 2016

More data has arrived during the last week or so and I have added these datasets to the collection.

This has raised the number of respondents to research group DI to 16 and to research group ND also to 16 which is at least balanced!

Where this is having an impact on the initial analysis is that it is enabling me to refresh the means and boundaries markers described above through use of the recently received data.

For research group DI, this has shifted the mean Dyslexia Index to 673.0 with a Standard Error of 35.77. This shifts the + or – 3 standard-error boundary points to 566.57 – 781.22.

My interpretation of this is that we would expect almost all students with dyslexia to present a Dyslexia Index in the range $567 < DI < 781$ and a Dyslexia Index that falls outside these boundaries might be regarded as a highly unusual result.

It is the lower boundary that is particularly useful in the data analysis of this project because it can be used as the criteria that marks a suspicion that a student with no prior indication of dyslexia who presents a Dyslexia Index above this boundary point *is* indicating a profile that is more in line with dyslexia than not. Hence, students from research group ND exhibiting this characteristic will form research DNI which is the one I am hunting for.

UPDATE 17th March 2016

Data still trickles in with a total of 19 datasets in research group ND and 18 in research group DI. It is at least pleasing that the numbers of dataset in each base research group remain similar.

As I include new datasets into the collective data spreadsheet, this re-adjusts the means and boundary points accordingly. With such small research groups I need to very carefully determine where boundary points are set and this remains a bit of a ‘dark art’ at present. I am planning for wider publicity for the research QNR next month as at the current time (mid March) university terms are ending and so the audience for e-mail invitations or the new poster campaign that I have designed is understandably limited. So the Big Push will be mid-April with a modest target of 100 datasets in total – surely this is possible?

Now: to return to the update here which is to keep track of the adjustments I am applying to the boundary points as further datasets arrive. I have to persist in attending to these re-calculations as this development process is key to ensuring that a robust application of statistical principles is applied to this initial stage of the data analysis.

The latest development has been to reconsider the application of the Central Limit Theorem to the datasets in the research group DI as it will be the mean and boundary points established here that will determine which datasets in the base research group ND are removed to create the key research group, DNI.

I believe that my application of the Central Limit Theorem to my stats is appropriate because I am working with samples – small ones at present – and I need to work with estimates for population parameters based on sample data, which is what the Central Limit Theorem is all about.

BlogPost #17: QNR DATA IS ARRIVING – FIRST IMPRESSIONS

We know that a 95% Confidence Interval for a population mean can be estimated from the sample mean +/- 1.96 standard errors. A 99% Confidence Interval is generated using alternative critical values of +/- 2.58 standard errors. So by working from the sample mean Dyslexia Index for research group DI – as this changes due to further datasets arriving – I am regularly updating the 99% CI for the population mean. Once these boundary values for the DI are established, I am then using the datasets from QNR respondents whose Dyslexia Indices fall within this 99% Confidence Interval for the population mean DI as the one from which I create the mean values for all the other parameters I am interested in. That is, for Academic Behavioural Confidence, and for the 6 psychometrics that I have previously referred to as constituting the Locus of Control Profiles (although I am now in favour of dropping the ‘Locus of Control’ bit – just an unnecessary and perhaps slightly misleading application of the term. However I may revise this later).

So with the latest dataset added to research group DI, new figures for the mean Dyslexia Index and for the 99% Confidence Interval have worked out at: sample mean Dyslexia Index: 602.94 with the 99% Confidence Interval for the population mean Dyslexia Index as:

$602.94 < \text{population mean DysInd} < 757.22$. These boundary points are now calculated using a Standard Error for the sample of 29.90.

On this basis, this identifies a core group of datasets in research group DI the Dyslexia Indices of which fall within this 99% CI range and at present, it is these datasets that are used to calculate the mean values for the other parameters I am exploring and which construct the base profile for this research group.

Now this idea transfers across to the datasets in research group ND where first of all I am applying the lower, 99% CI boundary point for Dyslexia Index as the determining criteria for shifting a dataset into research group DNI. That is, a determiner that a respondent who indicates no learning challenges on their QNR does, in fact, have a profile that is more aligned with those respondents who have declared dyslexia as their learning challenge, than with other profiles in the base research group ND.

UPDATE 21st April 2016

I now have 87 QNR replies of which 10 are ‘spoiled ballots’ mostly because the data is so incomplete that I can’t include it.

As data has arrived, the main QNR spreadsheet has been updated and mean averages for the 8 data scales recalculated. Profiles have been built and mean average data points on the profiles accordingly updated with recalculated scale averages.

In order to record these processes for write up later, just as I have recorded earlier settings above, I have settled on the following settings and protocols for now:

- Dyslexia Index is referred to as Dx and the Dx scale ranges from 0 to 1000
- a boundary setting of Dx = 500 is applied to differentiate between dyslexic profiles ($Dx > 500$) and non-dyslexic profiles ($Dx < 500$). This has been established not because 500 marks the midpoint of the scale but because from the data QNR replies forming research group DI, the mean average Dyslexia Index has settled at Dx = 668 with only 2 QNR respondents data generating Dx values of less than 500 (Dx = 444, Dx = 467). The standard error of the mean for this dataset is 23.57 and using standard Central Limit Theorem principles, this standard error value generates a lower 99% confidence limit of Dx = 607 for this research group. Hence it is felt that setting the boundary of Dx = 500 means that a QNR response indicating Dx < 500 is very highly unlikely indeed to be indicating that the respondent has a dyslexic learning profile – that is, assuming that this QNR respondent is *also* indicating that they have no learning challenges.
- Hence for research group ND – that is, students who are declaring no learning challenges – I am confident that this QNR respondents whose Dx > 607 may be shifted to research group DNI – that is, students with unidentified dyslexia. This is, of course, where the primary research interest in the whole project lies.

However it is noteworthy to remark at this stage that I am disappointed that the profile differentiation that I had hoped to see has not been readily apparent – that is, using profiles to differentiate between students with identified dyslexia and those with unidentified dyslexia. However the Dyslexia Index scale (Dx) that I developed based on responses to my earlier enquiry to dyslexia support professional across the HE sector in the UK appears to be working very effectively in that it corroborates the self-disclosure provided on QNR replies in research group DI. This gives confidence that it is effectively identifying students with a dyslexic learning difference profile and hence, it has been possible to easily spot students in research group ND – that is, with no disclosed dyslexic learning difference – who do, in fact, present a dyslexic learning profile. This is not to say that the profiles that all of the QNRs are generating are worthless but that I will be needing to review their contribution to the project. What remains clear is that there is a good deal of information locked up in the data that generates the profiles and exploring the meaning of all this will be a major focus.

UPDATE 4th May 2016

Data is trickling in very slowly despite repeated efforts to publicize the project and the research QNR. The most recent effort has been an e-mail request to the Student Support or Dyslexia/Disability services of most of the Higher Education institutions in the UK (128) where I asked that the link to the research QNR be promoted through an e-mail request inviting participants, circulated through student e-mail distribution lists. A few days later and of the acknowledgements received from this e-mail ‘hit’, all but one of the few that did reply refused to help. It is possible that the one university that agreed to publicize the project in their student newsletter will produce a few more QNR replies and publicity through posters is still up at both my ‘home’ research university and my employment university.

However, I now have 28 respondents who have disclosed their dyslexic learning differences on their QNR reply and a useful, 74 respondents who indicate no learning challenges or learning challenges other than dyslexia making 103 replies that have provided usable data. There have been no further ‘spoiled ballots’ to add to the existing 10 that are so incomplete as to provide unusable data.

Cursory inspection of this data is beginning to indicate that this hundred-or-so replies may be sufficient to enable some decent data analysis to take place.

In summary to date:

- In research group DI, the mean Dyslexia Index is settling at approximately $Dx = 693$ with only 3 of the 28 datasets indicating $Dx < 500$, which in itself is interesting and this anomaly will be explored in more detail later. The Academic Behavioural Confidence of research group DI is settling at approximately $ABC = 59$.
- Of the other 75 datasets nominally assigned as research group ND, 50 have $Dx < 500$ with a mean of $Dx = 362$ and $ABC = 69$.
- The remaining 25 datasets the Dyslexia Index is $Dx > 500$ so these are now re-assigned to form research group DNI and these datasets are generating a mean $Dx = 632$ and mean $ABC = 64$.

So on the basis of just eyeballing these means, it does, indeed look like students who are claiming no dyslexic learning differences but who are nevertheless presenting a dyslexic profile **are** indicating a higher Academic Behavioural Confidence than their identified, dyslexic peers.

UPDATE 20th May 2016

This is the final update of this post. A fresh StudyBlog entry has been created to outline data analysis decisions that emerging and to present a broad summary of results as they stand to date.

However it is worth reporting here that as the (revised) closing date for the data collection process is drawing close (31st May 2016) it has been very encouraging to report that a recent flurry of new QNR replies has been received. I think this is resulting from the ‘final push’ to recruit participants both at my home university and also through the e-mail circular I sent to other institutions requesting co-operation in alerting their students to this project and encouraging them to take part.

I now have a pool of 180 datasets in total of which 163 are providing good quality data. The remaining 17 were ‘spoiled’ in some way that makes the limited data that they contain not worth including. In most cases, this has arisen because respondents did not complete the questionnaire nor even partially complete it to the extent where data provided could usefully contribute to the project.

Of the 163 datasets that I can use, 68 are from students who are indicating dyslexia. I believe this to be sufficient to enable a baseline reference to be created for the 8 scales that I am measuring for students with dyslexia against which I can gauge other students’ responses.

Of the 95 datasets remaining, 73 are indicating *no learning challenges* with the remaining 22 comprising students disclosing learning challenges other than dyslexia (16) or *other learning challenges* that they are then not specifying.

I think that I have now collected sufficient data for the analysis that I am planning to generate meaningful results which is the next stage of the project.